

# HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis

Year: 2020 | Citations: 2369 | Authors: Jungil Kong, Jaehyeon Kim, Jaekyoung Bae

---

## Abstract

Several recent work on speech synthesis have employed generative adversarial networks (GANs) to produce raw waveforms. Although such methods improve the sampling efficiency and memory usage, their sample quality has not yet reached that of autoregressive and flow-based generative models. In this work, we propose HiFi-GAN, which achieves both efficient and high-fidelity speech synthesis. As speech audio consists of sinusoidal signals with various periods, we demonstrate that modeling periodic patterns of an audio is crucial for enhancing sample quality. A subjective human evaluation (mean opinion score, MOS) of a single speaker dataset indicates that our proposed method demonstrates similarity to human quality while generating 22.05 kHz high-fidelity audio 167.9 times faster than real-time on a single V100 GPU. We further show the generality of HiFi-GAN to the mel-spectrogram inversion of unseen speakers and end-to-end speech synthesis. Finally, a small footprint version of HiFi-GAN generates samples 13.4 times faster than real-time on CPU with comparable quality to an autoregressive counterpart.